IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 16, 2022, accepted February 17, 2022, date of publication February 25, 2022, date of current version March 4, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3154767

Predict the Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques

MUSTAFA A. AL-ASADI AND SAKIR TASDEMİR
Department of Computer Engineering, Faculty of Technology, Selçuk University, 42130 Konya, Turkey
Corresponding author: Mustafa A. Al-Asadi (masadi@lisansustu.selcuk.edu.tr)

ABSTRACT Football is a popular sport; however, it is a big business as well. From a managerial perspective, the important decisions that team managers make —Concerning player transfers, issues related to player valuation, especially the determination of transfer fees and market values, are of major concern. Market values can be understood as estimates of transfer fees— prices that could be paid for a player on the football market. Therefore, market values play an important role in transfer negotiations. The market has traditionally been estimated by football experts. However, expert judgments are inaccurate and not transparent. Data analytics may thus provide a sound alternative or a complementary approach to experts-based estimations of market value. In this study, we propose an objective quantitative method to determine football players' market values. The method is based on the application of machine learning algorithms to the performance data of football players. The data used in the experiment are FIFA 20 video game data, collected from sofifa.com. We estimate players' market values using four regression models that were tested on the full set of features—linear regression, multiple linear regression, decision trees, and random forests. Moreover, we seek to analyze the data and identify the most important factors affecting the determination of the market value. In the experimental results, random forest performed better than other algorithms for predicting the players' market values. It has achieved the highest accuracy score and lowest error ratio compared to baseline. The results show that our methods are capable to address this task efficiently, surpassing the performance reported in previous works. Finally, we believe our results can play an important role in the negotiations that take place between football clubs and a player's agents. This model can be used as a baseline to simplify the negotiation process and estimate a player's market value in an objective quantitative way.

INDEX TERMS Player value prediction, regression, machine learning, football analytics, FIFA video game data.

I. INTRODUCTION

Football is the world's most popular game in the number of participants and spectators [1]. The revenue for European football clubs alone for 2017 was rated at $27bn [2]. Therefore, it becomes an essential contributor to the global economy [3]. The demand for football stars has been increasing dramatically in the past few decades, and the value of the football players is exceeding €100 M. These numbers are much higher than historical trade figures, compared to the normal inflation rate [4]. From a management perspective, the most important decision that football clubs have to make is choosing the players [5]. Player transfers have a tremendous impact on a club's chances of success [6]. Therefore, researchers from various disciplines have studied the factors affecting transfer fees [4].

Recently, researchers have begun to pay special attention to the market values of players. A player's market value is an estimate of the amount with which a team could sell a player's contract to another team [7]. While transfer fees represent the actual prices paid in the market, market values provide estimates of transfer fees, therefore they play an important role in transfer negotiations. Football experts, such as team managers and sports journalists have always valued market values, while crowdsourcing sites such as Transfermarkt (www.transfermarkt.com) have proven useful in estimating

The associate editor coordinating the review of this manuscript and approving it for publication was Ehab Elsayed Elattar.

market value over the past few years. However, data-based methods for estimating market value have not been used widely in football [8].

The problem of identifying the most important determinants of market value for football players has been well described in the literature [9]–[12]. In the literature, many different variables were found and these indicators are classified into three categories—player characteristics, player performance, and player popularity. Some studies indicated that there is a non-linear relationship for some of these variables—such as age—with the dependent variable (market value) [7], [8], [13].

For the past two decades, machine learning has become essential to transforming football statistics into useful information for helping teams and coaches analyze opponents and make better decisions in real-time. However, research in football analytics with Machine Learning techniques is limited. The main reason for this is the lack of a large-scale dataset for players, which is a problem because gathering such detailed information about players may be expensive, making sensed data limited to teams with high purchasing power [1]. In football analytics, video games like FIFA and Football Manager (FM) consider as another source of data. Since 2014, researchers and clubs have used video games as alternative sources for data. Shin and Robert used the FIFA video game data to predict the result of the matches. They found that this data can be used in machine learning projects to make predictions with very accurate results [14]. In this paper, an effective machine learning method is introduced, designed using FIFA 20 dataset. This dataset includes different performance ratings for more than 17,000 players. Attributes of this dataset show different skills of players—shooting score, passing score, and dribbling score. Using this dataset, we can evaluate the performances of the players in the past season.

Theoretical hypothesis: To our knowledge, linear regression models have been used without regard to the fact that some variables are non-linearly related to the value of the player. This means that the use of nonlinear regression methods (such as decision trees) may show outperformance over the standard approach used in the literature.

The aim of the study was achieved using three steps:

1) The first step determined the factors affecting the players' market value and organized them into a dependent variable (i.e. market value) and independent variables (i.e. predictors).

2)The second step analyzed the selected factors that influence the market value of the players. This analysis was conducted in two stages:

- The first stage is a preliminary analysis to study the quality of the selected features.

- The second stage is an extended analysis in which the logical choice of these features is verified.

3) In the third step, in contrast to other methods used in the literature, linear and non-linear methods were tested to solve the problem.

The main novelty of this research is the study and analysis of the factors that affect the market value of football players across several stages and the estimation of the value of players based on their relevant features. The experimental results showed the superiority of the proposed non-linear methods over the latest methods in the problem of predicting the market value of football players. Thus, the contributions of this study are not limited to the field of application related to video games but go beyond it through the superiority of the methodology used in the study over the standard approach used in the literature to solve the same problem and using the same data.

Our results show that using random forests to predict the market value of football players is very promising. Compared to previous work Behravan and Razavi [13], we achieve much better accuracy by using a single learning model. Moreover, the random forest is easier to measure than the optimization method Behravan and Razavi [13] and the training of the model is relatively fast. In addition, the random forest model requires fewer inputs (7 inputs) than the Behravan and Razavi [13] model (55 inputs) and Müller *et al.* [8] model.

Sections 2, 3, 4, 5, 6, 7, 8, and 9 proceed with the background, the methodology of the study, dataset description, and the machine learning models used in the study, the evaluation metrics, the results, and the discussion respectively. The last section of the paper encompasses the conclusions and further research.

II. BACKGROUND

A. FACTORS THAT INFLUENCE THE MARKET VALUE OF FOOTBALL PLAYERS

Since 2013, the International Center for Sports Studies (CIES) has developed a robust econometric approach to assess the transfer value of professional footballers on a scientific basis [15]. According to this approach and previous literature, in this experiment, we looked at the different factors that influence the market value of football players. According to the literature, the most common indicators for assessing market value fall into three categories: player characteristics, player performance, and player popularity. In the next section, we review the most important selected studies that used these indicators.

1) PLAYER CHARACTERISTICS

Player characteristics are described as both physical and demographic attributes. ***Age*** is an important indicator of market value, as it reflects both experience and ability [16]. Most studies used the age factor to estimate market value, bearing in mind that players' values usually increase until their mid-20s and decrease thereafter. Besides, it has been found that ***player height*** leads to a significant increase in salary returns [17]. because it indicates good header ability that may increase the likelihood of scoring or preventing a goal [18]. Another characteristic that has been studied in player-valuation research is ***footedness***. Bryson *et al.* [17]

concluded that being able to play with both feet raises the salaries of players, and Herm *et al.* [7] found that it positively impacts their market values.

In the same context, the researchers also studied whether the players' ***nationalities*** affected their market values [4]. For example, in their study of the Spanish professional football league, Garcia-del-Barrio and Pujol [19] found that non-Spanish European players were systematically overrated, while non- European players were systematically underrated. Finally, the ***player position***—goalkeeper, defender, midfielder or forward player—is important in estimating market value. Several researchers have found that player positions affect salaries and transfer fees, as they reflect a player's degree of specialization and their ability to attract fans. Miao [20] concluded that attackers receive much higher attention and rewards than goalkeepers, as the attackers are more visible to the crowd and thus have a greater capacity to attract crowds.

2) PLAYER PERFORMANCE

Several player performance metrics can be used to estimate market values. ***Goals***, including field goals, headers and penalties, refer to players' ability to score and so are a largely unambiguous measure of performance [16].

Apart from the abovementioned metric, many researchers used other performance metrics that helped explain the value and the fees. ***Passing*** are used frequently [7]; ***duelling (or tackles)*** in the form of clearances; ***dribbles*** [21]; ***committed fouls*** [20]; and ***yellow and red cards*** [22].

3) PLAYER POPULARITY

In football, not only is the talent of the player crucial in determining the market value. The ***popularity*** also can explain the demand for football players [23]. In other words, the market value of football players also depends on their crowd-pulling power, independent of what they show on the pitch. The image of a player outside the football pitch influences the number of jerseys sold and money earned from portrait rights. Accordingly, studies of the football transfer market have investigated popularity-related factors [24]. Popular athletes have commercial value, which is important for the club [25]. Even though players like Messi, Ronaldo or even Ibrahimovic are close to retirement, their brand value is still very high as they have gained international stature during their careers. Everyone knows their face, and this gives them extra ammunition when negotiating sponsor deals with popular brands. In summary, this study has identified several indicators of market value, including player characteristics, performance, and popularity, with most of the extant studies relying on similar factors. The next section explains how we operationalized these factors and how we analyzed the dataset to train market-value estimation models.

B. RELATED WORK

Bhravan and Razavi built a machine learning model using the FIFA 20 dataset. In their study, they used Hybrid regression—a combination of particle swarm optimization (PSO) and support vector regression SVR). According to the authors, the RMSE and MAE for their method are 2,819,286 and 711,029.413, respectively, while the results presented by [8] were 5,793,474 and 3,241,733. These results indicate that their method has a significant advantage over other methods of estimating the market value of football players [13].

Philippi *et al.* analyzed the impact of team variables and player positions on the market value of football players. According to their results, the regression analysis showed that team level, birth month, league, place of play, and player's age influence the players' market values. They also indicated that players who play in attacking midfield and were born in the first quarter of the year are the most valuable players [26].

Müller *et al.* [8] presented a multi-level regression method for estimating the market value of players. They created a dataset that contained various attributes such as player characteristics—age, position, nationality—player performance, and popularity. They analyzed the influence of various factors on the market value of players and then trained a regression method to estimate the value. Besides, the authors in their paper explained the limitations of the crowdsource estimating method used by transfermarkt.com.

Majewski [12] investigated the influence of various factors on the value of forwarding players to determine the most important factors. In this study, he used information on 150 famous attackers stored in Transfemarkt.de and adopted the GLS method (generalized at least squared) to find the important factors. Based on his results, the number of goals, assists, the value of the entire team, and FIFA's rating points had an impact on the market value of attacking players. Considering the role of players as the positive point in his study, but merely focusing on the forward players is a negative point.

Stanojevic and Gyarmati [27] presented a methodology for estimating the market value of 12,858 players based on players' performance data. Where they built several models using supervised learning and players' performance data gathered from transfermarkt.uk website and sports analytics company InStat. These models were built using 45 predictors. The results proved that the developed model outperforms widely used transfermarkt.com market value estimates in predicting team performance.

Herm *et al.* [7] introduced a method to estimate the transfer fee of football players based on five talent variables (age, precision, success, assertion, and flexibility). Their model shows the age is inversely proportional to players' market value. So, the main drawback of this research is using community evaluations that can be biased or suffer from a lack of knowledge.

Frank and Noisch [23] investigated the impact of talent and popularity of players on their market value. They attempted to measure players' talent using 20 criteria. Using an OLS regression model, they concluded that the popularity of players increases their market value.

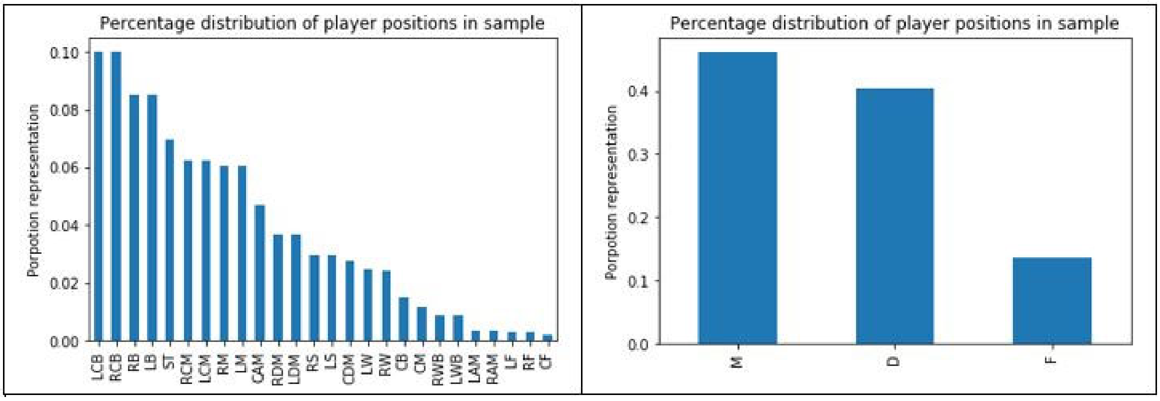


FIGURE 1. Distribution of players positions in the original dataset before and after grouped.

III. METHODOLOGY

The methodology used in this experiment is based on supervised machine learning. That is, the algorithm "learns" from samples of data to infer a model, and then the model is tested using other samples that were not used to build the model. These test samples allow us to compare the values predicted by the model with actual values, and gauge the model's accuracy in predicting real samples. In this study, the expected values are the market value of football players, and each sample is represented by a set of variables that represent the player's performance and skills. The proposed methodology consists of the following steps:

• Step 1: General investigation to determine the factors affecting the players market value:

In this step, we have reviewed the studies devoted to predicting a player's market value. Besides, we searched the literature for the factors affecting the market value, and we identify the variables that have an impact on the market value of players. Nine variables were identified due to their frequent appearance in the literature. Table 1 shows the features that were used in this study.

• Step 2: Preprocessing techniques:

Preprocessing is one of the most data mining tasks which includes preparation and transformation of data into a suitable form to mining procedure. It includes several techniques like data cleaning, transformation, reduction, etc [28]. Data cleaning is a very important step in any machine learning project to get accurate results. Thus, the data has been cleaned and processed, and the most appropriate part of the data was used while building the models.

The data cleaning steps are summarized as follows:

- Removing the redundant columns (e.g. name, link, id etc.) and keeping columns with the features we want for modelling as dependent variables (including Age, Height, Potential, International reputation, Weak foot, Team position, Shooting, Passing, and Dribbling).

- Dealing with the missing values.
- Converting categorical features (Team Position) to numeric values.

There are 26 unique locations in the data set. This is too much for a categorical variable. Therefore, these positions have been grouped into more general categories as in Figure1 (e.g. forward, midfielder, and defender).

- Converting all numeric columns to integer or float.
- Converting International Reputation column (5 ⋆) to integer (5) with scale (1-5).
- Converting Weak foot column (5 ⋆) to integer (5) with scale (1-5).
- Converting player value to decimal money (Including €, M and K characters).
- Applying a log transformation on the data to reduce the variance in the target variable.

The player value is our target feature. The values of different players vary significantly, particularly for the top players whose value increases exponentially. Analyzing this data showed us that the values of players have a right-skewed distribution which could make the prediction of very high-value players difficult during modelling. The logarithmic transformation has been applied to give it a better distribution (Figure. 2).

• Step 3: The preliminary analysis of a selected subset of features

To study the quality of the selected subset of features (identified in step 1), the level of interdependence of these features with each other was studied using **the Pearson Correlation Coefficient**. The hypothesis on which the heuristic is base states below: Good feature subsets contain features highly correlated (predictive of) with the class, yet uncorrelated with (not predictive of) each other. Figure 3 summarizes the numerical features correlations to the target variable (Player value). Besides this, the predictors (features) themselves were also correlated with each other. This can be seen in Figure 4.

As shown in Figures 3 and 4, playing attributes (shooting, passing and dribbling) tend to be correlated strongly with each other but do not strongly correlate to player value. Therefore, it may be necessary to combine some of these attributes to reduce the model complexity. Of the playing attributes, the passing seems to be the most correlated to

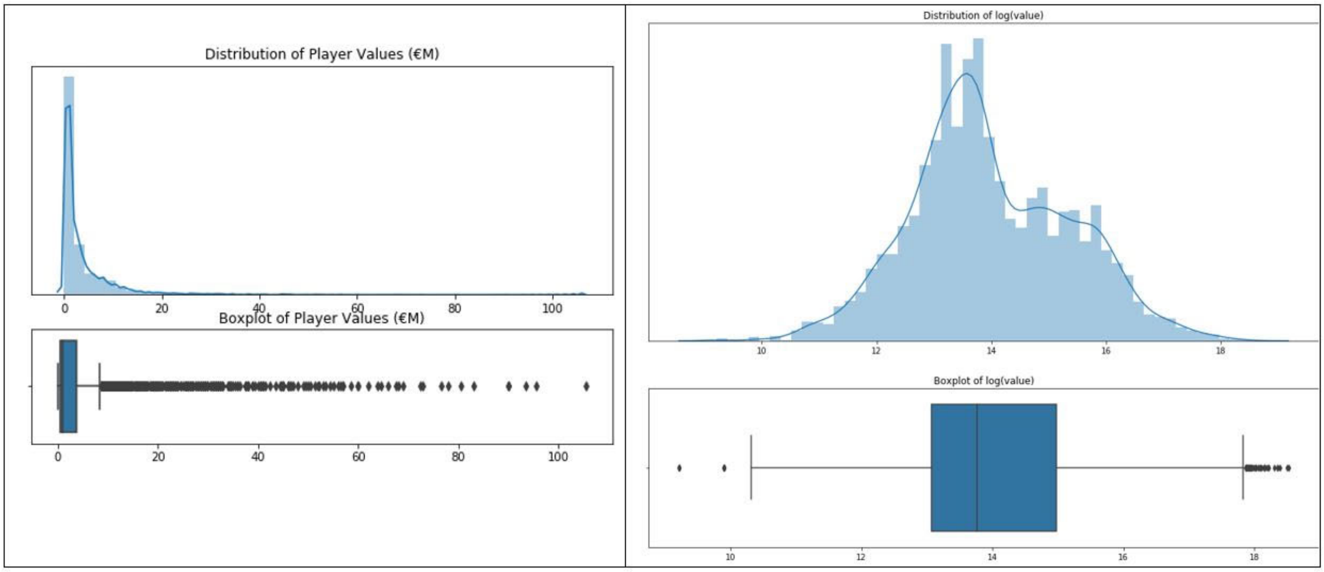


FIGURE 2. Football player value distribution before and after logarithmic transformation.

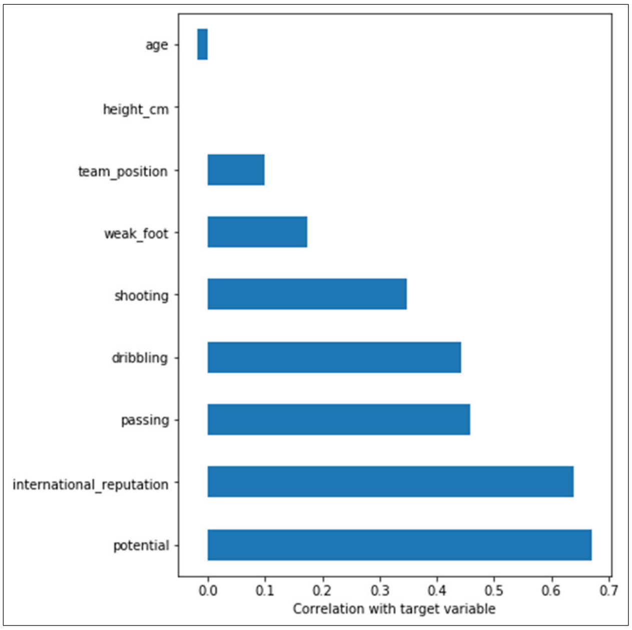


FIGURE 3. The numerical features correlation to the target variable (Player value).

the value. Therefore, we decided to consider the passing attribute and dropping shooting and dribbling. Figure 5 shows the correlation matrix for selected attributes in the term of Heatmap.

- **Step 4: The extended analysis of a selected subset of features**

After defining the new subset of features (identified in Step 3), the logical selection of these variables was verified by statistical significance using linear regression analysis and decision trees. In the experiments, an ordinary least squares (OLS) model was fitted. Analysis of variance could provide us with a first impression of how each predictor was correlated with the dependent variable. The P-value was used to determine the statistical significance of the regression coefficients. P-value allows telling whether the null hypothesis is to be rejected or not.

Figure 6 shows **the Analysis of Variance (ANOVA)** table via OLS fit. According to the analysis, the coefficients of age, height, potential and international reputation are significant (P_value < 0.05). As for the variables weak foot, position, and passing, they are not statistically significant. This indicates that these features increase model complexity without improving performance and should be considered candidates for dropping in the multiple linear regression model.

For decision tree methods, **Gini Importance** calculates the importance of each feature as the sum over the number of splits (across all trees) that include the feature, proportionally to the number of samples it splits [29]. Figure 7 shows the importance of predictors according to models of decision trees and random forests. On the left in Figure 7, the four most important variables in determining a player's value are potential, age, international reputation, and height. On the right in Figure 5, the four most important variables in determining a player's value are potential, age, international reputation, and passing skill. According to this analysis, passing skill is of greater importance than the player's height, while linear regression reveals that passing skill is not statistically significant.

- **Step 5: Data splitting**

After cleaning the dataset and defining the new subset of features and verifying the logical selection of these variables, 80% of the data has randomly allocated to train the classifier, and the remaining 20% was used for testing.

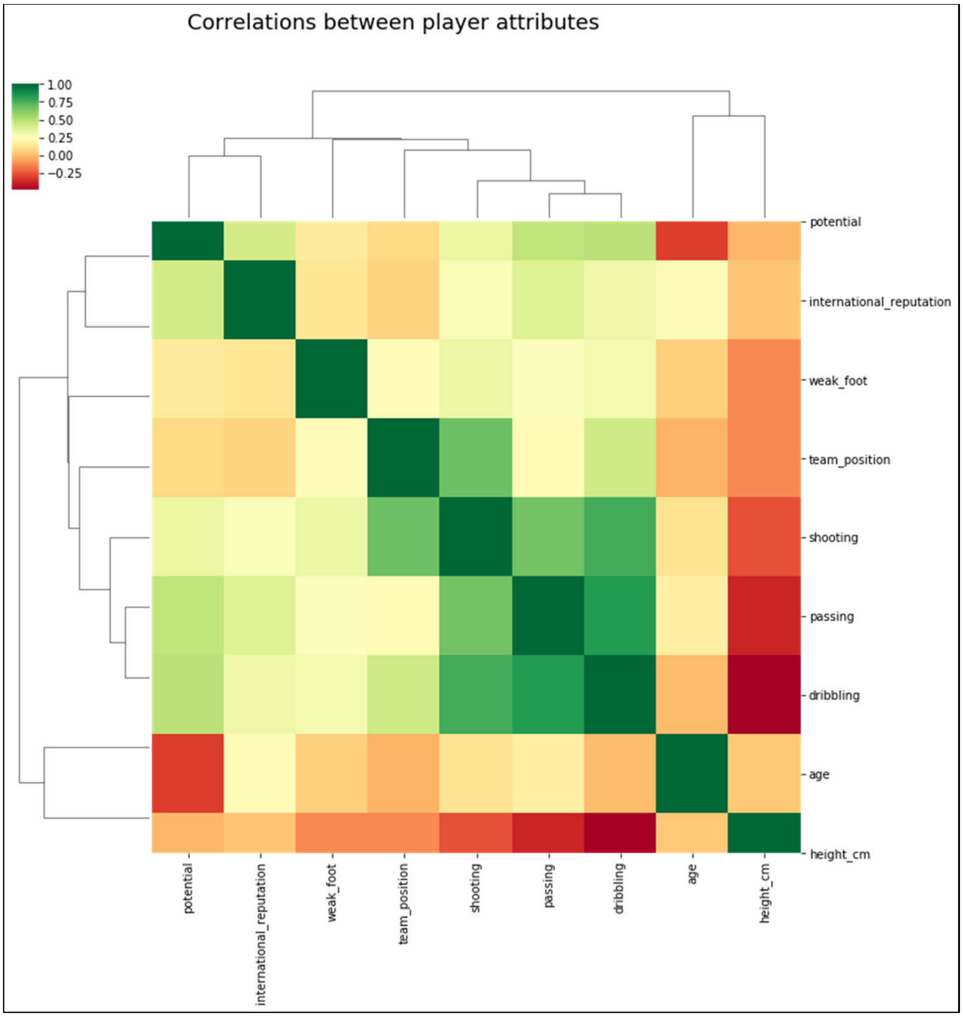


FIGURE 4. Heatmap for origin attributes.

• Step 6: Modelling the market value models

In the modelling process, the player market value was used as an objective function of 17980 players. We estimated players' market values using four regression models that were tested on the full set of features (linear regression, multiple linear regression, decision trees, and random forests). All models were created using the default parameters unless otherwise noted. The results of the models were compared to the real values, and it is found that it is applicable for this aim.

Step 7: Evaluation models:

To evaluate the performance of the models, several metrics were calculated like, Train and Test Split is used to estimate model performance using the training set. Mean absolute errors (MAE), Root mean square errors (RMSE) and The coefficient of determination (R^2) is used to evaluate the regression models using the testing data. a Python module called scikit-learn is used to build machine learning models.

IV. DATASET DESCRIPTION

The difficulty in obtaining large-scale reliable data and the cost problems related to this process were explained in the Introduction section. For these reasons, in this study, we aim to use the FIFA soccer video game data, which is commonly used in the literature. It has been used successfully to predict the results of football matches [30], and we have seen that it was comparable or better than other sources of football data [14]. Therefore, we believe that the results of the video game data set can be correlated with market transactions of real football players and other analytics.

The EA Sports FIFA video game series system began in 2009. It offers detailed information, including weekly updates, about a broad set of European soccer players and their skills, which covers three aspects: physical, mental, and technical skills. This information is available on the official website of the game (http://sofifa.com/).

In football analytics, video games like FIFA and Football Manager (FM) consider as another source of data. Since

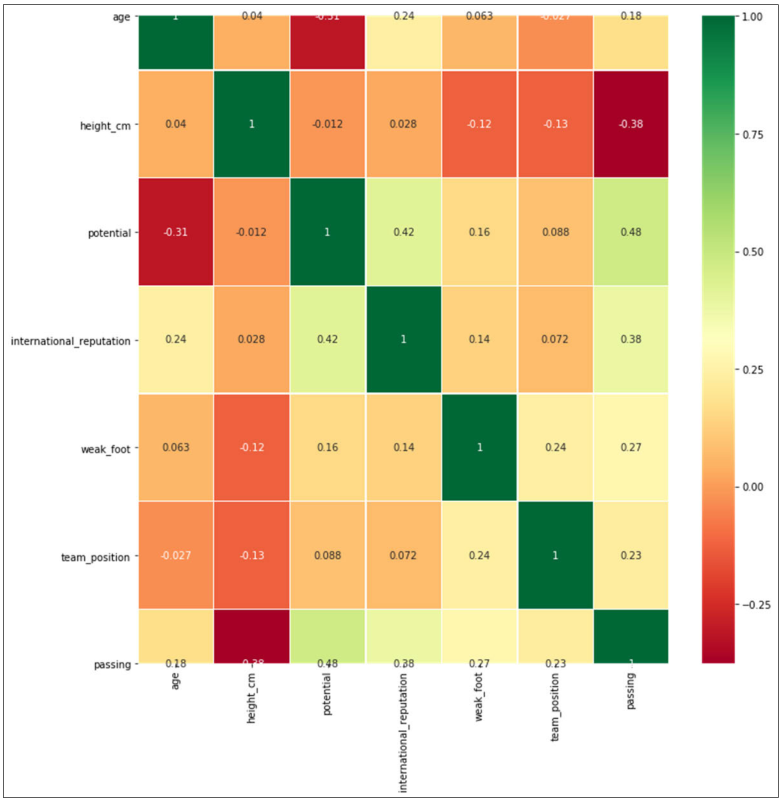


FIGURE 5. Heatmap for selected attributes.

2014, researchers and clubs have used video games as alternative sources for data.

Football Manager is a sports simulation game for team management, developed by Sports Interactive company. In 2014, Sports Interactive company signed a deal with Prozone, to use data from Sports Interactive in their online software called Prozone Recruiter. Many of the top clubs to scout new players use prozone Recruiter software. Premier League clubs have begun using the database belonging to Football Manager to help identify and recruit new signings since 2014 [31], [32].

EA Sports employs a wide network of real-life scouts who attend matches and watch tapes of games to determine skill scores for each player in the game. These can change between each installment of the series. For instance, if a young player gets better at shooting or dribbling during a real-life season, his score will be increased for the next edition of FIFA. Likewise, if an older player starts to slow or loses stamina, this will also be reflected in his score. EA Sports employs an extensive network of real-life scouts who attend matches and watch game streams to determine skill scores for each player in the game. A score for each of these criteria is assigned to every player, and these are used to help calculate an overall rating out of 100. Each player in the game has over 300 fields as well as over 35 specific attributes which ultimately determine the rating seen in the game [33], [34]. However, there is no formula or scientific equation for FIFA to determine the market value of a player. Scouts generally do the job using their experience and player scores, which exposes the determination of market value to many biases.

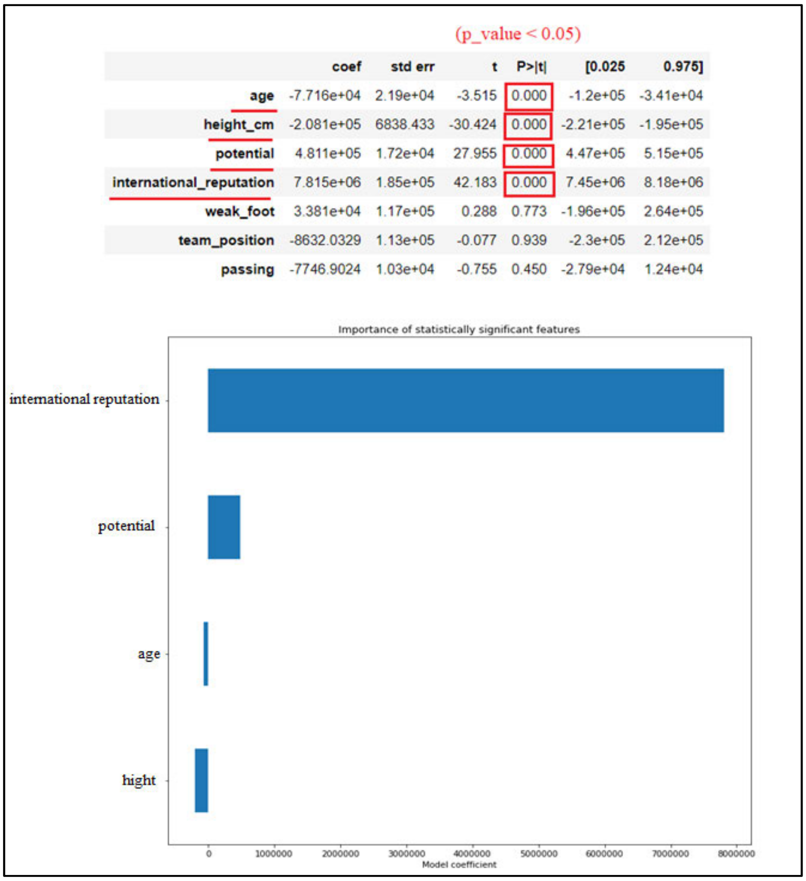
Recently, Leone [35] compiled, cleaned, and shared a dataset of statistics of European professional football. He used the EA Sports' FIFA 20 video game series system for organizing an Excel database. This data allows finding insights about the footballers' performance from a quantitative perspective. Further, these data were successfully used by Awasthi *et al.* [36] in their study.

In this study, we used the FIFA 20 dataset, provided and shared by Leone on Kaggle.[1] It contains 17,980 cases, and each case is about one football player. Each football player has more than 70 attributes. These attributes can be divided into personal attributes (e.g., age, nationality, and value), performance attributes (e.g., overall, potential, and stamina), and value. For our analysis, we selected nine continuous variables (as a dependent variable) and a player's value (as an independent variable).

V. MACHINE LEARNING ALGORITHMS

To predict the player market value from variables that reflect the skills and characteristics of a football player, four different supervised machine learning methods were used. All of these

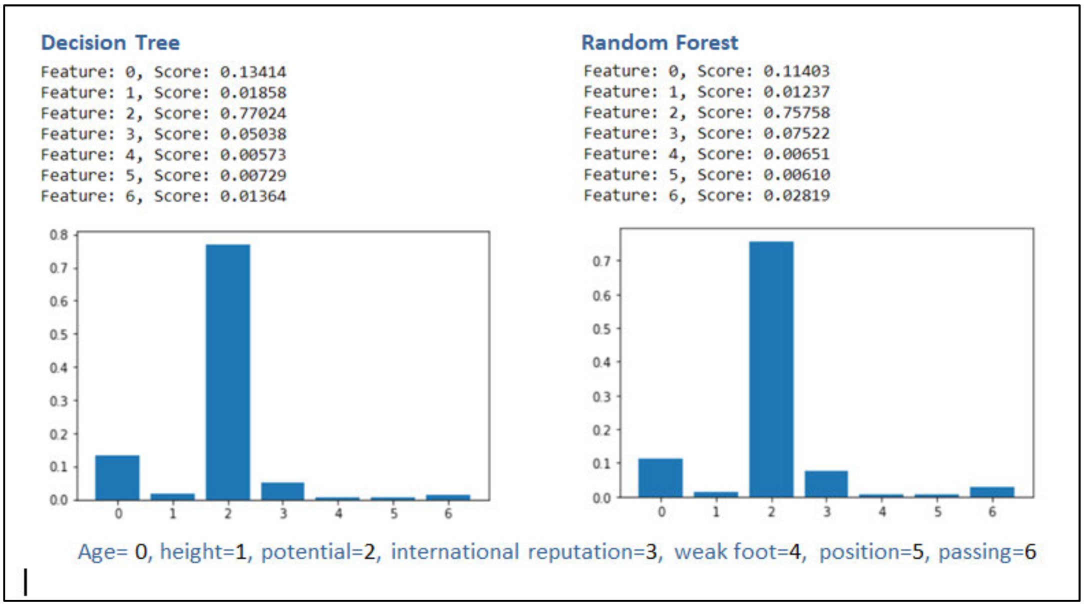
[1] https://www.kaggle.com/stefanoleone992/fifa-20-complete-player-dataset



(p_value < 0.05)

	coef	std err	t	P>|t|	[0.025	0.975]
age	-7.716e+04	2.19e+04	-3.515	0.000	-1.2e+05	-3.41e+04
height_cm	-2.081e+05	6838.433	-30.424	0.000	-2.21e+05	-1.95e+05
potential	4.811e+05	1.72e+04	27.955	0.000	4.47e+05	5.15e+05
international_reputation	7.815e+06	1.85e+05	42.183	0.000	7.45e+06	8.18e+06
weak_foot	3.381e+04	1.17e+05	0.288	0.773	-1.96e+05	2.64e+05
team_position	-8632.0329	1.13e+05	-0.077	0.939	-2.3e+05	2.12e+05
passing	-7746.9024	1.03e+04	-0.755	0.450	-2.79e+04	1.24e+04

FIGURE 6. ANOVA table via OLS fit.



Decision Tree

```
Feature: 0, Score: 0.13414
Feature: 1, Score: 0.01858
Feature: 2, Score: 0.77024
Feature: 3, Score: 0.05038
Feature: 4, Score: 0.00573
Feature: 5, Score: 0.00729
Feature: 6, Score: 0.01364
```

Random Forest

```
Feature: 0, Score: 0.11403
Feature: 1, Score: 0.01237
Feature: 2, Score: 0.75758
Feature: 3, Score: 0.07522
Feature: 4, Score: 0.00651
Feature: 5, Score: 0.00610
Feature: 6, Score: 0.02819
```

Age= 0, height=1, potential=2, international reputation=3, weak foot=4, position=5, passing=6

FIGURE 7. The importance of predictors according to models of decision trees and random forests.

methods aim at providing the optimal link between patterns of skills and the market value of the players. Because the market value is a numerical variable, all methods that were selected can predict continuous numerical values.

TABLE 1. Variables used for the market value analysis.

N	Variable	Possible value	Description
1	Age	-	Age reflects players' experience and potential.
2	Height_cm	-	Height reflects heading ability, which can influence the probability of scoring or preventing goals.
3	Potential	1-100	The potential is how much a player can grow during his career mode save.
4	International_reputation	1, 2, 3, 4 or 5	is an attribute that affects the player's rating according to his club's local and international prestige.
5	Weak_foot	1, 2, 3, 4 or 5	Two-footedness is an advantageous footballing ability that also reflects players' flexibility.
6	Team_position	1, 2, or 3	Position reflects players' flexibility and their role on the pitch (eg forward, midfielder, and defender).
7	Shooting	1-100	The player's ability to shoot the ball with strength
8	Passing	1-100	Passing refers to the number of passes to other players or the accuracy of passing.
9	Dribbling	1-100	Dribbling refers to the number and success rate of a player's ball manoeuvres.

The supervised machine learning methods used in this experiment cover four machine learning paradigms, and include Linear Regression, Multiple Linear Regression, Regression Tree, and Random Forest Regression. These algorithms have been chosen according to their frequent use in the literature of characterizing players [37], [38] and data mining domains. Moreover, they are relatively fast state-of-the-art algorithms [39], [40]. Additionally, these algorithms were selected for performance comparison between nonlinear methodologies (such as decision trees) and linear methodologies (such as linear regression).

In statistics, linear regression is a linear approach to modelling the relationship between a numerical response and one or more explanatory variable(s) (also known as dependent and independent variables).

A line will be created in the multiple linear regression by determining the coefficients.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \epsilon \tag{1}$$

The best-fitted line to the dataset will be the optimum combination of coefficients that minimizes the residual sum of squares (RSS).

$$RSS = \sum\nolimits_{i=1}^{n} \left(y_i - \check{\beta}_0 - \check{\beta}_1 X_{i1} - \check{\beta}_2 X_{i2} \ldots - \check{\beta}_p X_{ip}\right)^2 \tag{2}$$

The formula below is used in decision tree-based models to minimize RSS value [41].

$$RSS = \sum\nolimits_{j=i}^{j} \sum\nolimits_{i \epsilon R_j} \left(y_i - \check{y}_{Ri}\right)^2 \tag{3}$$

Decision tree learning is one of the predictive modelling techniques used in machine learning. It uses a decision tree—as a predictive model—to move from observations about an element—represented in branches—to conclusions about the element's target value—represented in the papers. Decision trees where the target variable can take continuous values—usually real numbers—are called regression trees. Decision trees are among the most popular machine learning algorithms due to their clarity and simplicity [42].

Random forests are an ensemble learning method for classification, regression, and other tasks that work by creating multiple decision trees at training time and outputting a class that represents the category placement (classification) or average prediction (regression) of individual trees. Random forests generally outperform decision trees. However, data characteristics can affect its performance [43], [44].

VI. EVALUATION METRICS

To evaluate the performance of the models, several metrics were calculated. For example, the Train and Test Split is used to estimate model performance using the training set. Mean absolute errors (MAE), Root mean square errors (RMSE), and the coefficient of determination (R^2) is used to evaluate the regression models using the testing data. In machine learning, player market value can be handled in different ways. We can consider it a regression problem and expect the market value based on the data of players' performance. In this study, we have established four regression models. The data of players' performance and skills was used as features in building models—to build the baseline and compare results.

A. TRAIN AND TEST SPLIT

The simplest way to evaluate the algorithm's performance is to use different sets of training and testing. In this technique, the original data is split into two parts. The first part trains the algorithm and makes predictions on the second part and then evaluates predictions against the expected results. Generally, the size of the split data is based on the size of the dataset. The common use is 70-80% for the training and 20-30% for testing [45]. In our study, the data were randomly divided into 70% for training and 30% for testing.

B. ERROR MEASUREMENTS

Each machine learning model is trying to solve a problem with various objects using different data. Usually, in regression problems Mean absolute errors (MAE), Mean square errors (MSE), Root mean square errors (RMSE), and the coefficient of determination (R^2) is used for evaluating the

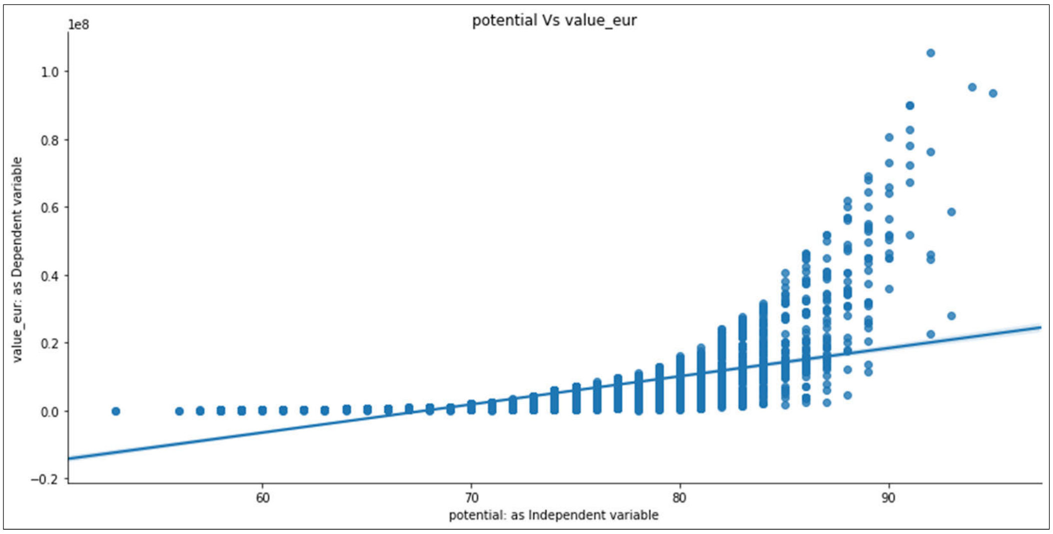


FIGURE 8. Scatter chart and simple linear regression (potential vs value).

model, as formalized in Equations (4) to (6).

$$MAE = \frac{1}{N}\sum\nolimits_{i=1}^{N}\left|y_i - \check{y}_i\right| \tag{4}$$

$$MSE = \frac{1}{N}\sum\nolimits_{i=1}^{N}\left(y_i - \check{y}_i\right)^2 \tag{5}$$

$$\mathrm{R}^2 = \frac{1 - MSE(Model)}{MSE(Baseline)} \tag{6}$$

where y_i is the actual expected output and $\check{y}_i$ is the model's prediction.

Mean absolute error (MAE) measures the average error size in a set of predictions, regardless of their direction. Mean square errors (MSE); measure the average squared error of predictions. It measures the square difference between the predictions and the target and then the mean of those values for each point. The higher this value, the worse the model is. It is never negative but would be zero for a perfect model. Root mean square errors (RMSE) is just the square root of MSE. The square root is introduced to make the errors' scale the same as the targets' scale [46]. The coefficient of determination (R^2) is another metric we may use to evaluate a model, and it is closely related to the MSE of the model and baseline. MSE baseline is the simplest possible model would get. The simplest viable model would always be to predict the average of all samples. When evaluating the model, a value close to one indicates a model with close to zero error, and a value close to zero indicates a model very close to the baseline [47].

VII. RESULTS

By applying the methods described above to the data described in the data set desecration section—using the different machine learning algorithms—provided the results as shown below:

A. BASELINE MODEL (LINEAR REGRESSION)

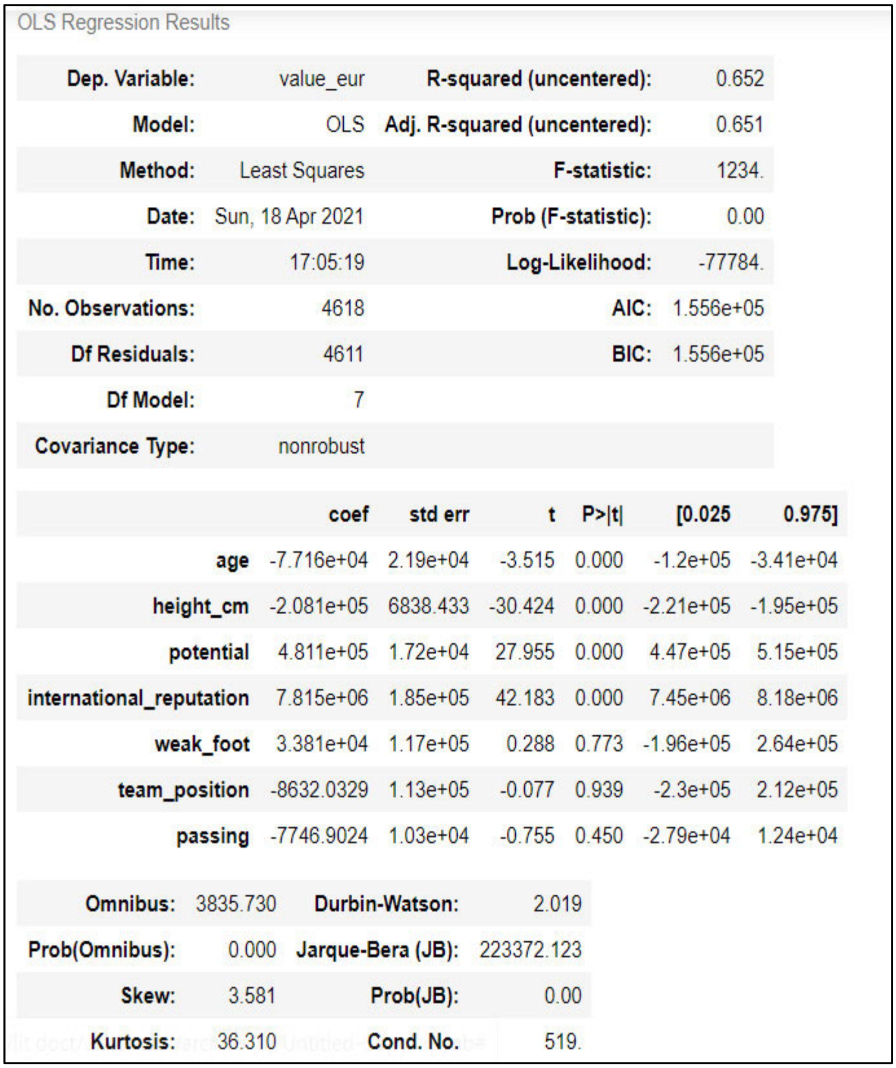
A baseline is a bedrock for a machine learning model's lowest acceptable performance on a premier dataset. Generally, suppose a model achieves a performance below the baseline. In that case, it will be a failure, and we should try a different model or admit that using machine learning techniques to improve the model is not right for our problem. From the methodology section (step 3), we discovered that the potential of the player had the highest correlation with the value of the player (0.66). Therefore, the simplest model with good performance is likely to be a linear regression of potential rating versus the log (market value).

The baseline linear regression gives an RMSE of 5.46 and an R-squared score of 0.43. This is not a good model because there is still room for improvement. Moreover, it does not give any insight into other variables which could affect the value of each player. Figure 8 shows the linear regression results when using player potential only.

B. MULTIPLE LINEAR REGRESSION

The multiple linear regression model is an improvement on the baseline model, with an RMSE of 4.66 and an R-squared score of 0.56. The adjusted r2 is also higher than the simple linear regression, indicating that although the model is more complex, the added complexity improves the predictive performance of the model. Figure 9 shows the multiple linear regression results when using all the available features.

From the EDA analysis, it appears that the international reputation, potential, height, and age of players have high absolute feature importance as expected. The player position and most player attributes are statistically insignificant for predicting player value. This indicates that these features increase the complexity of the model without improving performance and should be neglected when modelling multiple linear regression.



OLS Regression Results

Dep. Variable:	value_eur	R-squared (uncentered):	0.652
Model:	OLS	Adj. R-squared (uncentered):	0.651
Method:	Least Squares	F-statistic:	1234.
Date:	Sun, 18 Apr 2021	Prob (F-statistic):	0.00
Time:	17:05:19	Log-Likelihood:	-77784.
No. Observations:	4618	AIC:	1.556e+05
Df Residuals:	4611	BIC:	1.556e+05
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
age	-7.716e+04	2.19e+04	-3.515	0.000	-1.2e+05	-3.41e+04
height_cm	-2.081e+05	6838.433	-30.424	0.000	-2.21e+05	-1.95e+05
potential	4.811e+05	1.72e+04	27.955	0.000	4.47e+05	5.15e+05
international_reputation	7.815e+06	1.85e+05	42.183	0.000	7.45e+06	8.18e+06
weak_foot	3.381e+04	1.17e+05	0.288	0.773	-1.96e+05	2.64e+05
team_position	-8632.0329	1.13e+05	-0.077	0.939	-2.3e+05	2.12e+05
passing	-7746.9024	1.03e+04	-0.755	0.450	-2.79e+04	1.24e+04

Omnibus:	3835.730	Durbin-Watson:	2.019
Prob(Omnibus):	0.000	Jarque-Bera (JB):	223372.123
Skew:	3.581	Prob(JB):	0.00
Kurtosis:	36.310	Cond. No.	519.

FIGURE 9. Multiple linear regression results.

C. REGRESSION TREE

Regression trees provided a significant improvement over the baseline model, with an RMSE of 2.71 and an R-squared score of 0.87. This means that the regression trees are a powerful generalization of the linear regression algorithm. Regression trees are known to do a better job at capturing nonlinearity in data by dividing the space into smaller subspaces, depending on the questions asked.

According to our dataset, there is a nonlinear relationship between the value of the player and their age and height, as shown in Figure 3. This explains why regression trees outperform linear regression. As illustrated in Figure 10, the algorithm starts with all data at the root node and you apply the linear regression formulas there for every bipartition of every feature. For each feature, the algorithm calculates the MSE (Mean squared error) per sample for every possible partition along the feature axis. This approach can create more complex decision boundaries, unlike a single straight line, which is not flexible enough.

D. RANDOM FOREST REGRESSION

Random forest is an ensemble of decision trees. Many trees, constructed in a certain 'random' way, form a Random Forest. Random Forest Regression also provided a significant improvement over the baseline model, with an RMSE of 1.64 and an R-squared score of 0.95. This means that the Random Forest Regression is a powerful generalization of the linear regression algorithm and for Regression trees.

Random Forest Regression, like Regression trees, also does a better job at capturing nonlinearity in data by dividing the space into smaller subspaces, depending on the questions asked. As illustrated in Figure 11, each tree is created from a different sample of rows and at each node; a different sample of features is selected for splitting. Then, each of the trees makes its prediction. Finally, these predictions are then averaged to produce a single result.

Figure 12 shows the mean absolute error between the predicted and actual values, when using the different machine learning algorithms. As the figure shows, the Random Forest algorithm provided the lowest mean absolute difference

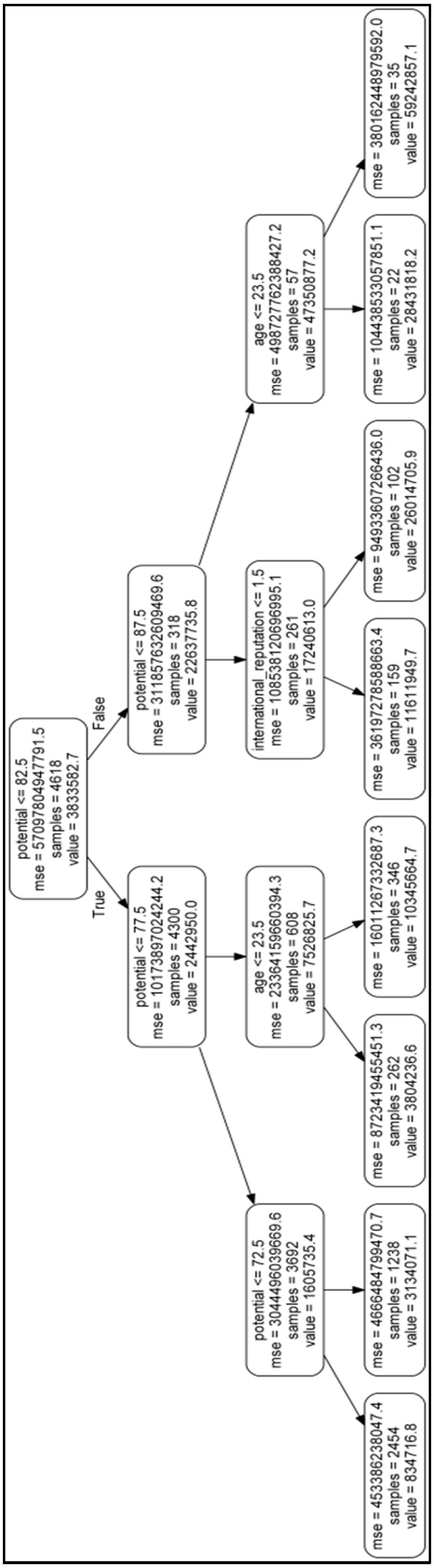


FIGURE 10. Regression Trees flowchart diagram used for player market value prediction with min samples leaf = 1 and max depth = 3.

between the predicted and actual value, Linear Regression provided the highest mean absolute error.

TABLE 2. Shows mean absolute errors (MAE), Root mean square errors (RMSE) and the coefficient of determination (R^2) for all models.

N	Classifier	MAE	RMSE	R²
1	Linear Regression (Baseline)	5,468,144	5,468,144	0.47
2	Multiple Linear Regression	2,618,108	4,662,630	0.61
3	Regression Tree	835,935	2,713,452	0.87
4	Random Forest Regression	576,874	1,649,921	0.95

Figure 13 shows the root mean square errors between the predicted and actual values when using the different machine learning algorithms. The Random Forest algorithm provided the lowest root mean square errors difference between the predicted and actual value; Linear Regression provided the highest root mean square errors.

Figure 14 shows the coefficient of determination (R^2) for each machine learning algorithm used. A value close to 1 indicates a model with close to zero error. In contrast, a value close to zero indicates a model very close to the baseline. As the figure shows, the Random Forest algorithm provided the highest value for the coefficient of determination (R^2), Linear Regression provided the lowest value for the coefficient of determination (R^2). This means that the Random Forest algorithm is the best for modelling.

VIII. DISCUSSION

A player's market value is an estimate of the amount with which a team could sell their contract to another team [7]. Therefore, it plays an important role in the negotiations that take place between football clubs and the player's agents. In this paper, we adopted a quantitative method based on machine learning that depends on the skills and performance of the player and other factors. We also sought to compare linear and nonlinear methods according to the data set used. According to the data analysis, the football player market value is affected by numerous factors that are not directly related to performance or skills. For example, a player's international reputation is the second most important feature in determining a player's value, after the player's potential, according to all the methods used by the experience. Therefore, we conclude that the market value of the players depends to a large extent on their crowd-pulling power, regardless of their performance on the pitch.

Our results show that using random forests to predict the market value of football players is very promising. Compared to previous work [13], we achieve much better accuracy by using a single learning model (see Table 3). Moreover, the random forest is easier to measure than the optimization method in [13] and the training of the model is relatively fast. In addition, the random forest model requires fewer inputs (7 inputs) than the [13] model (55 inputs) and [8] model. We also assure that the Random forest model not only has excellent performance but also can more accurately calculate the significance of the variables. Therefore, it was encouraging that

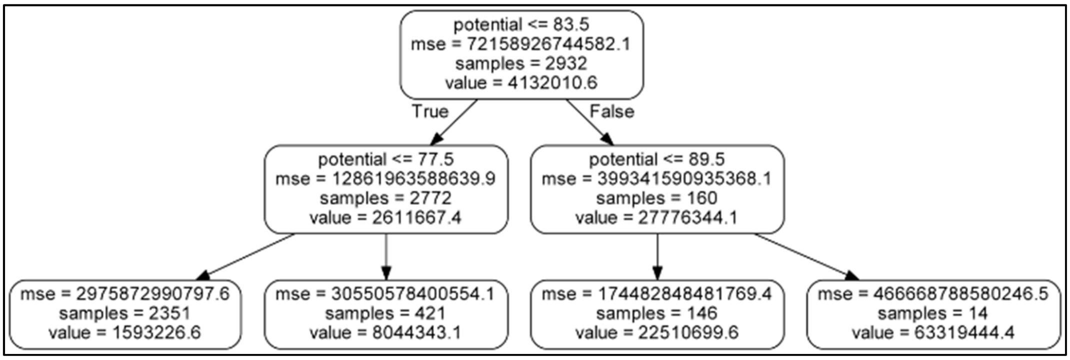


FIGURE 11. Random Forest Regression flowchart diagram used for player market value prediction with the number of estimators = 10 and max depth = 3.

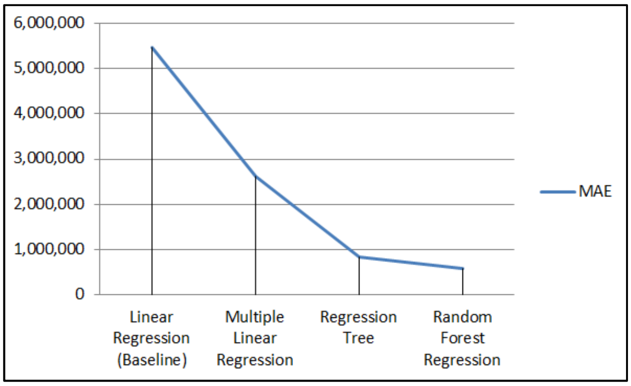


FIGURE 12. Mean absolute error when using Train and test split scheme.

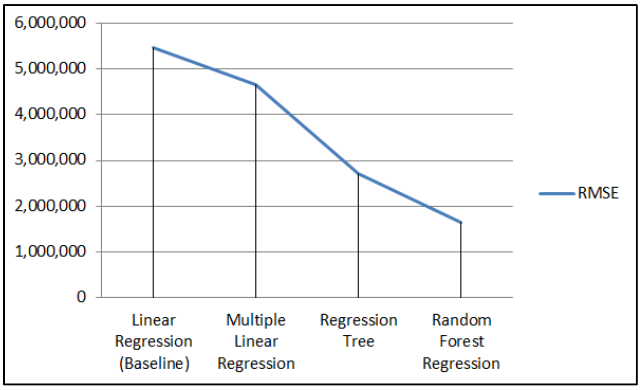


FIGURE 13. Root mean square errors when using Train and test split scheme.

our modelling and analysis technique managed to do better, in fact, way better.

In theoretical implications, our study assumes that the use of nonlinear regression methods (such as decision trees) may show outperformance over the standard approach used in the literature. Our results confirm the superiority of random forest over other linear models. Thus, our findings are consistent with the theory on which the research is based, and this has been confirmed.

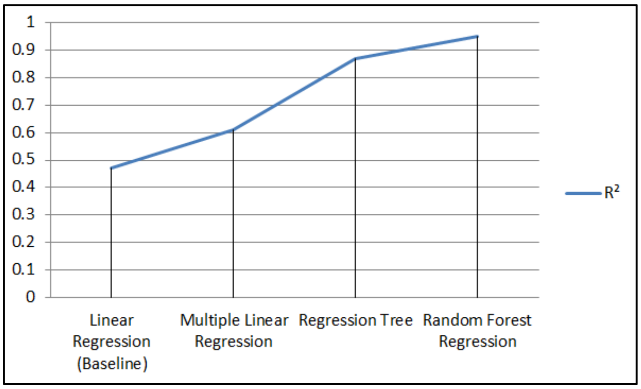


FIGURE 14. Coefficient of determination (R^2) for each machine learning algorithm.

TABLE 3. Evaluation of the studies in the literature.

Study No.	Algorithm Used	MAE	RMSE	R²	No. features
[13]	A combination of PSO and SVR	711,029	2,819,286	0.74	55
[8]	Multi-level regression	3,241,733	5,793,474	-	22
This study	LR (Baseline)	5,468,144	5,468,144	0.47	7
	MLR	2,618,108	4,662,630	0.61	
	RT	835,935	2,713,452	0.87	
	RF	**576,874**	**1,649,921**	**0.95**	

It wouldn't be wrong to say that decision tree models were more reliable due to the high ranking success in replicating training data on the testing data. Moreover, decision tree models have the advantage of their visualization like a flowchart diagram and there are no assumptions about distribution because of the nonparametric nature of the algorithm [48].

IX. CONCLUSION AND FURTHER RESEARCH

The ability of video games to simulate football has progressed rapidly in the past two decades [49]. Besides, great efforts have been invested in analyzing the skills and performance of soccer players to allow reliable simulations of games that reflect the realities of soccer matches. FIFA datasets showed effectiveness in predicting football match results and other analyses. The results show it was equal or better than other football data [14], [50], [51].

The experimental results of our study showed the superiority of the proposed non-linear methods over the latest methods in the problem of predicting the market value of football players. Thus, the contributions of this study are not limited to the field of application related to video games but go beyond it through the superiority of the methodology used in the study over the standard approach used in the literature to solve the same problem and using the same data.

In the context of FIFA games, FIFA Ultimate Team (FUT) is a game mode in FIFA that allows players to build and manage their club using different cards to play offline or online matches [52]. In FUT, the player must spend virtual currency to get packs. When you start playing Ultimate Team in FIFA, the goal is to build a team made up of the best players and cards which are often very expensive. Lots of people buy FIFA Points with real money until they reach their goal to unlock special packs, but spending on virtual players in the game is not the right solution [53]. Therefore, the best solution is to trade team players to earn money, develop and improve the team. FIFA 20 automatically sets a standard price for each player in the game, but video players don't trust this value too much. It is, therefore, necessary to objectively determine the price of the player and to know whether the price of that player will rise or fall before making any buying or selling in the market. Thus, this will allow determining the average value at which the player is currently sold. Then you can use this method to set the player price to help get as many coins as possible. In future studies, the results of this study can be used to create a calculator hosted on the FIFA website to aid video gaming players and generate a financial gain.

Finally, we believe our results can play an important role in the negotiations that take place between football clubs and a player's agents. In conclusion, these models can be used as a baseline to simplify the negotiation process and estimate a player's market value in an objective quantitative way.

CONFLICT OF INTEREST STATEMENT

Mustafa A. Al-Asadi declares that the submission is an original study that is not under review by any other journal on behalf of all authors. There are no financial conflicts of interest to disclose.

REFERENCES

[1] L. Cotta, "Using fifa soccer video game data for soccer analytics," in *Proc. Workshop Large Scale Sports Anal.*, 2016, pp. 1–4.

[2] R. Vroonen, "Predicting the potential of professional soccer players," in *Proc. Mach. Learn. Data Mining Sports Anal. ECML/PKDD Workshop*, 2017, pp. 1–10.

[3] R. Asif, "Football (soccer) analytics: A case study on the availability and limitations of data for football analytics research," *Int. J. Comput. Sci. Inf. Secur.*, vol. 14, no. 11, p. 516, 2016.

[4] B. Frick, "The football players' labor market: Empirical evidence from the major European leagues," *Scottish J. Political Economy*, vol. 54, no. 3, pp. 422–446, 2007.

[5] E. Amir and G. Livne, "Accounting, valuation and duration of football player contracts," *J. Bus. Finance Accounting*, vol. 32, nos. 3–4, pp. 549–586, 2005.

[6] T. Pawlowski and C. A. Breuer Hovemann, "Top clubs' performance and the competitive situation in European domestic football competitions," *J. Sports Econ.*, vol. 11, no. 2, pp. 186–202, 2010.

[7] S. Herm and H.-M. H. Callsen-Bracker Kreis, "When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community," *Sport Manage. Rev.*, vol. 17, no. 4, pp. 484–492, 2014.

[8] O. Müller, A. Simons, and M. Weinmann, "Beyond crowd judgments: Data-driven estimation of market value in association football," *Eur. J. Oper. Res.*, vol. 263, no. 2, pp. 611–624, Dec. 2017.

[9] P. Wicker *et al.*, "No pain, no gain? Effort and productivity in professional soccer," *Int. J. Sport Finance*, vol. 8, no. 2, pp. 124–139, 2013.

[10] T. Idson and L. Kahane, "Team effects on compensation: An application to salary determination in the national hockey league," *Econ. Inquiry*, vol. 38, no. 2, pp. 345–357, Apr. 2000.

[11] L. M. Kahn, "The sports business as a labor market laboratory," in *The Business of Sports*, S. R. Rosner and K. L. Shropshire, Eds. Sudbury, MA, USA: Jones and Bartlett, 2004, pp. 242–251.

[12] S. Majewski and U. Szczecin, "Identification of factors determining market value of the most valuable football players," *J. Manage. Bus. Administration. Central Eur.*, vol. 24, no. 3, pp. 91–104, Sep. 2016.

[13] I. Behravan and S. M. Razavi, "A novel machine learning method for estimating football players' value in the transfer market," *Soft Comput.*, vol. 25, no. 3, pp. 2499–2511, 2021.

[14] J. Shin and R. Gasparyan, "A novel way to soccer match prediction," Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, 2014, pp. 1–5.

[15] R. Poli, L. Ravenel, and R. Besson, "CIES football observatory," *Annu. Rev.*, pp. 1–83, 2014. [Online]. Available: https://www.football-observatory.com/IMG/pdf/ar2013_exc-2.pdf

[16] F. Carmichael and D. Thomas, "Bargaining in the transfer market: Theory and evidence," *Appl. Econ.*, vol. 25, no. 12, pp. 1467–1476, Dec. 1993.

[17] A. Bryson, B. Frick, and R. Simmons, "The returns to scarce talent: Footedness and player remuneration in European soccer," *J. Sports Econ.*, vol. 14, no. 6, pp. 606–628, 2013.

[18] T. R. L. Fry, G. Galanos, and A. Posso, "Let's get Messi? Top-scorer productivity in the European champions league," *Scottish J. Political Economy*, vol. 61, no. 3, pp. 261–279, 2014.

[19] P. Garcia-del-Barrio and F. Pujol, "Hidden monopsony rents in winner-take-all markets—Sport and economic contribution of Spanish soccer players," *Managerial Decis. Econ.*, vol. 28, no. 1, pp. 57–70, 2007.

[20] H. Miao and R. A. Cachucho Knobbe, "Football player's performance and market value," in *Proc. 2nd Workshop Sports Anal., Eur. Conf. Mach. Learn. Princ. Pract. Knowl. Discovery Databases (ECML PKDD)*, 2015, pp. 1–9.

[21] S. Medcalfe, "English league transfer prices: Is there a racial dimension? A re-examination with new data," *Appl. Econ. Lett.*, vol. 15, no. 11, pp. 865–867, Sep. 2008.

[22] S. Kiefer, "The impact of the Euro 2012 on popularity and market value of football players," Inst. Org. Econ., Berlin, Germany, 2012. [Online]. Available: https://www.wiwi.uni-muenster.de/io/forschen/downloads/DP-IO_07_2020

[23] E. Franck and S. Nüesch, "Talent and/or popularity: What does it take to be a superstar?" *Econ. Inquiry*, vol. 50, no. 1, pp. 202–216, Jan. 2012.

[24] J. Hofmann, O. Schnittka, M. Johnen, and P. Kottemann, "Talent or popularity: What drives market value and brand image for human brands?" *J. Bus. Res.*, vol. 124, pp. 748–758, Jan. 2021.

[25] A. Arai, Y. J. Ko, and S. Ross, "Branding athletes: Exploration and conceptualization of athlete brand image," *Sport Manage. Rev.*, vol. 17, no. 2, pp. 97–106, Apr. 2014.

[26] J. L. Felipe, A. Fernandez-Luna, P. Burillo, L. E. de la Riva, J. Sanchez-Sanchez, and J. Garcia-Unanue, "Money talks: Team variables and player positions that most influence the market value of professional male footballers in Europe," *Sustainability*, vol. 12, no. 9, p. 3709, May 2020.

[27] R. Stanojevic and L. Gyarmati, ''Towards data-driven football player assessment,'' in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2016, pp. 167–172.

[28] S. A. Alasadi and W. S. Bhaya, ''Review of data preprocessing techniques in data mining,'' *J. Eng. Appl. Sci.*, vol. 12, no. 16, pp. 4102–4107, 2017.

[29] K. Archer and R. V. Kimes, ''Empirical characterization of random forest variable importance measures,'' *Comput. Statist. Data Anal.*, vol. 52, no. 4, pp. 2249–2260, Jan. 2008.

[30] D. Prasetio, ''Predicting football match results with logistic regression,'' in *Proc. Int. Conf. Adv. Inform.: Concepts, Theory Appl. (ICAICTA)*, Apr. 2016, pp. 1–5.

[31] G. Crawford, D. Muriel, and S. Conway, ''A feel for the game: Exploring gaming 'experience' through the case of sports-themed video games,'' *Converg., Int. J. Res. New Media Technol.*, vol. 25, nos. 5–6, pp. 937–952, Dec. 2019.

[32] D. Beer, ''Productive measures: Culture and measurement in the context of everyday neoliberalism,'' *Big Data Soc.*, vol. 2, no. 1, 2015, Art. no. 2053951715578951.

[33] L. Yaldo and L. Shamir, ''Computational estimation of football player wages,'' *Int. J. Comput. Sci. Sport*, vol. 16, no. 1, pp. 18–38, Jul. 2017.

[34] R. Obiedat, ''Identification of players positions in a multi-agent game using artificial neural networks and C4.5 algorithm: A comparative study,'' *Sci. Res. Essays*, vol. 8, no. 17, pp. 682–688, 2013.

[35] S. Leone. (2020). *FIFA 20 Complete Player Dataset*. [Online]. Available: https://www.kaggle.com/stefanoleone992/fifa-20-complete-player-dataset

[36] P. Awasthi, A. Beutel, M. Kleindessner, J. Morgenstern, and X. Wang, ''Evaluating fairness of machine learning models under uncertain and incomplete information,'' in *Proc. ACM Conf. Fairness, Accountability, Transparency*, Mar. 2021, pp. 206–214.

[37] V. Rao and A. Shrivastava, ''Team strategizing using a machine learning approach,'' in *Proc. Int. Conf. Inventive Comput. Informat. (ICICI)*, Nov. 2017, pp. 1032–1035.

[38] K. Apostolou and C. Tjortjis, ''Sports analytics algorithms for performance prediction,'' in *Proc. 10th Int. Conf. Inf., Intell., Syst. Appl. (IISA)*, Jul. 2019, pp. 1–4.

[39] P. Manghi, L. Candela, and G. Silvello, *Digital Libraries: Supporting Open Science: 15th Italian Research Conference on Digital Libraries, IRCDL 2019, Pisa, Italy, January 31-February 1, 2019, Proceedings*, vol. 988. Cham, Switzerland: Springer, 2019.

[40] K. Gao, T. M. Khoshgoftaar, and A. Napolitano, ''Aggregating data sampling with feature subset selection to address skewed software defect data,'' *Int. J. Softw. Eng. Knowl. Eng.*, vol. 25, no. 09n10, pp. 1531–1550, 2015.

[41] J. Gareth, *An Introduction to Statistical Learning: With Applications in R*. Springer, 2013.

[42] X. Wu, V. Kumar, J. R. Quinlan, and J. Ghosh, ''Top 10 algorithms in data mining,'' *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.

[43] S. M. Piryonesi and T. E. El-Diraby, ''Role of data analytics in infrastructure asset management: Overcoming data size and quality problems,'' *J. Transp. Eng., B: Pavements*, vol. 146, no. 2, Jun. 2020, Art. no. 04020022.

[44] T. Kam Ho, ''Random decision forests,'' in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, Aug. 1995, pp. 278–282.

[45] J. Brownlee, *Machine Learning Mastery With Python*. San Juan, Puerto Rico: Machine Learning Mastery Pty Ltd, 2016, pp. 100–120.

[46] C. J. Willmott and K. Matsuura, ''Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,'' *Climate Res.*, vol. 30, no. 1, pp. 79–82, 2005.

[47] J. Dufour, ''Coefficients of determination,'' Dept. Econ., McGill Univ., Montreal, QC, Canada, 2011, pp. 1–14. [Online]. Available: https://monde.cirano.qc.ca/~dufourj/Web_Site/ResE/Dufour_1983_R2_W.pdf

[48] B. K. Gacar and I. D. Kocakoç, ''Regression analyses or decision trees?'' *Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, vol. 18, no. 4, pp. 251–260, 2020.

[49] A. S. Markovits and A. I. Green, ''*FIFA*, the video game: A major vehicle for soccer's popularization in the United States,'' *Sport Soc.*, vol. 20, nos. 5–6, pp. 716–734, 2017.

[50] D. Prasetio and D. Harlili, ''Predicting football match results with logistic regression,'' in *Proc. Int. Conf. Adv. Inform.: Concepts, Theory Appl. (ICAICTA)*, Aug. 2016, pp. 1–5.

[51] M. A. Al-Asadi and S. Tasdemir, ''Empirical comparisons for combining balancing and feature selection strategies for characterizing football players using FIFA video game system,'' *IEEE Access*, vol. 9, pp. 149266–149286, 2021.

[52] H. Wardle, ''When games and gambling collide: Modern examples and controversies,'' in *Games Without Frontiers?* Cham, Switzerland: Springer, 2021, pp. 35–77.

[53] P. Siuda, ''Sports gamers practices as a form of subversiveness–the example of the FIFA ultimate team,'' *Crit. Stud. Media Commun.*, vol. 38, no. 1, pp. 75–89, 2021.

MUSTAFA A. AL-ASADI received the master's degree in computer engineering from Selçuk University, Konya, Turkey, in 2018. He is currently pursuing the Ph.D. degree in computer engineering. His master's thesis focused on employing machine learning approaches to develop decision support systems for football team management. His research interests include machine learning, deep learning, predictive models, data analysis, data mining, and pattern recognition.

SAKIR TASDEMİR received the master's and Ph.D. degrees from Selçuk University, in 2004 and 2010, respectively. He is currently the Dean of the Faculty of Technology, Selçuk University, and the Head of the Computer Engineering Department. He has authored many publications in international journals. His research interests include decision support systems, expert systems, image processing, artificial intelligence, and computer aided systems. He is a member of an editorial board of several journals.